What You Use Matters: Coding Protest Data

Author(s): Taehyun Nam

Source: *PS: Political Science and Politics*, Apr., 2006, Vol. 39, No. 2 (Apr., 2006), pp. 281-287

Published by: American Political Science Association

Stable URL: https://www.jstor.org/stable/20451736

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

American Political Science Association is collaborating with JSTOR to digitize, preserve and extend access to *PS: Political Science and Politics*

This content downloaded from 34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

What You Use Matters: Coding Protest Data

Taehyun Nam, *Rhodes College*

Data matter. Protest data are no exception. Not all datasets are equal, nor do all have equal validity and reliability. Thus far, there has been no serious debate regarding the quality of protest datasets. Increasing attention to protest studies suggests that now is the time to talk about the *sine qua non* for growing knowledge in this sub-field of comparative politics. In this paper I emphasize the importance of the underlying data sources. In particular, I argue that protest datasets should be drawn from many sources, particularly as many local sources as possible, in order to provide accurate and meaningful data. To support this contention, I first discuss traditional protest datasets and their limitations and provide suggestions about how to overcome these limitations. Second, I explain an alternative type of dataset coding and its characteristics. And finally, I use statistical tests to illustrate the importance of local sources.

Limitations of Traditional Data

Event coding has become an essential part of data collection in various subfields of political science (Woolley 2000). Coding events, such as military conflicts,[1] often across countries and over a long period of time, has given scholars opportunities to conduct meaningful comparative analyses, thereby allowing them to avoid the potential bias associated with small-n studies (King, Keohane, and Verba 1994). Protest events, such as strikes or riots, are also coded and have become widely available. Some of the most commonly used protest event datasets include *Cross-National Time Series, 1815–1999* (Banks 2001) and *World Handbook of Political and Social Indicators* (Taylor and Jodice 1983), the contributions of which are indispensable to the study of protests. However, in spite of such contributions, these datasets have practical and theoretical limitations which have inhibited further development of protest research. The problems are associated with the generalization of categories of events, loss of information, and lack of adequate sources.

First, categories of events in the traditional datasets can obstruct efforts to discern their specific nature because their categorizations are often too general, arbitrary, and inaccurate. In these datasets, protest events are typically categorized into no more than a dozen types, according to multiple factors. These factors usually include the basic nature of an event (political/social), such as the level of violence, participants' characteristics and numbers, target, purpose, scope, and level of spontaneity. The end product typically gives the frequency of certain types of events for a given time period. This categorization provides a straightforward data structure. However, this practice produces several problems: 1) categories of protests are confusing because of coding inconsistencies; 2) an event may fall into multiple categories or no category at all; 3) the practice discards potentially valuable details; and 4) it limits the possibility for an end-user to sort and recode events based on his or her need.

For example, as Table 1 illustrates, the *World Handbook* has six categories for political protest—protest demonstration, regime-support demonstration, political strike, riot, armed attack, and political assassination (Taylor and Jodice 1983, 19–47). "Riot" is clearly an event distinguished from "demonstration" because of its violent nature, "destruction of property." However, unlike "demonstration" which is divided into anti-and pro-regime, the targets of riots are left unspecified, leading to the incorrect assumption that political riots occur only as protests against regimes.[2] Actually, only "demonstration" has been sub-categorized into anti and pro government. In addition, a riot, to be categorized as such, needs people in "hundreds, if not thousands" (29). Is a protest with all the attributes of "riot," but with only 90 participants, not a riot? If not, what is it? No answer is given. Again, this aspect of "riot" introduces the question of the number of participants in the other types of protest. How many participants are required for an event to be called a "demonstration"? Is this a meaningful criterion for other types of protest? If not, why is it the case that only "riot" has a minimum number of participants? Unfortunately, these questions are also left unanswered. Moreover, according to the data, there were two riots in South Korea in 1990. However, the available information about these riots goes no further than this. It remains unclear what the riots were about, who the targets were, or how violent the events were. This lack of specific information about events, in turn, deprives users of the ability to contemplate their attributes and recode events according to their interests. And it also precludes the users' ability to verify the reliability of data simply because one will not be able to find vital information pertaining to the coded event.

The second problem is that only yearly-summed numbers of events are available, eliminating the possibility of quarterly or monthly, let alone weekly or daily, analyses. For example, one finds from Banks' dataset (2001) that in South Korea there were six protest events—two riots and four anti-government protests—in 1990. However, no information is available on exactly when in 1990 these events occurred, where in South Korea, and what these events specifically entailed. Therefore, this annually aggregated dataset is insufficient if a researcher wanted to analyze protest events within a year[3] or if the main interest was the variation of protests on a daily basis over time.

Third, and most important, the traditional protest datasets have often relied on a limited number of sources. It is common for only one or two sources to be sampled. The *Prodat* project sampled German protest events from two German newspapers while Kriesi and his colleagues (1995) used only one national newspaper for each country.[4] To make matters worse, they both only coded events from Monday issues.[5] The ratio of sources to countries (*Prodat* project, 2:1; Kriesi et al., 1:1) significantly decreases in Banks' data (1991), *Domestic Conflict Behavior, 1919–1966*, which has only one source, the *New York Times*, for a total of 114 countries (1:114). It is hard to imagine, despite its fame for international coverage, that the *Times* is capable of covering all protests of a single country, let alone "virtually all nations

Taehyun Nam *is assistant professor, Department of International Studies at Rhodes College. His main interest includes the dynamics of protests and repression. The author can be contacted regarding this paper at namt@rhodes.edu. Data and codebook are available at his web site: www.poliscience.org.*

This content downloaded from
34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

Table 1
Categories of Political Protests and Their Factors in the *World Handbook of Political and Social Indicators* (Taylor and Jodice 1983)

	Categories of Event		
Factors	Protest Demonstration	Regime-Support Demonstration	Political Strike
Violence	No	No	
Actor	Organized people	Organized people	Industrial or service workers; students
Number of Participants			
Target	A regime or government, its leaders; ideology, (intended) policy, or lack of policy; previous action or intended action	A "government, its policies and actions, or one or more of its leaders"	"A regime and its leaders' policies or actions"
Purpose	"Altering the composition or policy of the government"	Lend support	To "embarrass the government, to erode its base of support, and even to precipitate its fall"
Event	Marching, hunger strikes, sit-ins, refusal to pay taxes, open letter, and individual self-destruction		A "work stoppage or a stoppage" of "normal academic life"
Scope		Issues of nationally important "all branches and levels of government"	Not "primarily directed at economic goals"
Planned			

	Categories of Event		
Factors	Riot	Armed Attack	Political Assassination
Violence	Destruction of property	Violent	Murder
Actor		Organized domestic group	
Number of Participants	"hundreds, if not thousands"		
Target		The "persons and property of politically significant groups or the representatives and facilities of the state itself"	A national leader, a high government official, or a politician
Purpose		"weakening or destroying the power exercised by another group"	Politically motivated
Event		Sabotage, terrorism, liberation, or kidnapping	
Scope		Political, excluding civil strife	
Planned	Spontaneity	Planned	

Source: Taylor and Jodice (1983, 19–38).

independent" (Banks 1972, 41). In terms of space, the *Times* has about a dozen sections. Sections like business, education, health, obituaries, science, technology, or weather are unlikely to report protests. So only a few sections remain. Among the remnants—front page, international, New York region, national, and Washington—only two sections, front page and international, are likely to report protests overseas. The space to report protests of the world is simply too small for full coverage. Therefore, when protests are sampled from one or two sources, especially non-local sources, the data is very unlikely to contain as many events as they intended or claimed to. Moreover, if potential coding bias—due to divergent trends of coverage among different media (Woolley 2000, 160–161) or changing indexing practices (162)—is considered, the risks involved in relying on a small number of sources significantly increase.

A New Perspective: Protest and Coercion Data

The cited problems can be overcome by 1) coding events with minimal generalization; 2) providing needed information about events and sources; and 3) increasing the number of sources. *Protest and Coercion Data* (PCD)[6] methods adopt all of these solutions at once. Hence, PCD provide users with data containing much more information about protest events in greater volume and quality. First, the problem of inconsistency of categorizations is resolved by focusing on an event itself rather than on a category. Each event is sampled as a distinctive incident regardless of its duration, number of participants, purpose, target, or property damage. Detailed information about each event is filed to form data. Thus, users have details about

This content downloaded from
34.192.2.131 fff:ffff:ffff on Thu, 01 Jan 1976 12:34:56 UTC
All use subject to https://about.jstor.org/terms

each event rather than getting summarized information about a collection of events. Recorded information of an event includes the event date, identity of protesters,[7] their target, the issue, and a brief description of the event. Activities of protest and coercion are informed by the number of protesters, the number injured, the number arrested, the number of deaths, state actors present (i.e., police), and the number of state actors injured and killed. Finally, the data include source information, such as the name of newspapers and publication date.

Such information is gathered and coded in PCD, creating detailed information about each event. Such an abundance of information allows the end-user to sort events into whichever categories he or she wishes. For instance, from the French PCD, we know that on Monday April 25, 1983, about 30 "farmers" stopped and emptied two trucks carrying Italian wines in Sete, France in protest against EU imports. In the course of the event, property-damage occurred, but no one was injured or arrested. According to the *World Handbook*, this event must be categorized somehow, but it is not clear in which category this event would belong. It could have been coded as "riot" because of the property damage, but there were only about 30 participants, which would disqualify this event as a riot according to the same codebook. Instead, it could be coded as an "armed attack" because this category does not specify the minimum number of participants and it does cover events with property damage. However, it is difficult to determine if the action was planned—which is an important criterion for an event to be "armed attack." The more serious problem is that once this event is coded, much of the information about the event disappears; instead it increases the count of a category by one. On the other hand, from the perspective of PCD users, this event does not need to be categorized, and various information of this event can be used in multiple ways. Or, if needed, this event can be categorized in more than one category. Since the information pertaining to the event is included in the event description, users can always go back, review the detailed information, and change the categorization to fit their coding. The farmers' behavior can be categorized as a riot, armed attack, demonstration, "foreign policy" (by topic), or "under 100" (by the number of protesters) depending on the needs of different researchers.

Second, to overcome the yearly aggregated data problem, the PCD project codes sub-daily events. As mentioned, PCD contain every and all reported protests and repressive events in a country. All events in a given day are separately coded, and an event is distinctly coded if it evolves into a different one. For example, assume that an event occurs where students protest on a campus. Later that day, the students try to march outside of the campus, and a riot breaks out when the police try to stop them. In terms of students' behavior, there are two distinct events. The protest on campus does not involve any confrontation, but it turns into a violent event against state actors. Here two events, rally and riot, are coded as separate events on the same day, and the numbers of participants in each event (some protesters do not take part in the riot) are also separately coded. This practice of coding allows researchers to study protests on a sub-annual basis. One can trace, for example, weekly or daily trends of protests if necessary, which gives a more detailed and informed knowledge about protests than studies based on yearly data. Thus, PCD allow researchers to find patterns that are almost impossible to study using annually aggregated data.

The third requirement, data from a small number of sources, is solved by simply employing many relevant sources. PCD have multiple sources. For example, the European PCD used approximately 500 newspapers, trade publications, and wire services available on Lexis-Nexis as their main sources. The magnitude of sources emphasized by PCD provides a critical improvement: events sampled in PCD outnumber other datasets by huge margins (Francisco 2006). For example, from 1980 through 1982, in France, 975 events were coded in PCD, while only 202 were coded in the *World Handbook* and none in Banks' data (1991). Even in comparison with data focused on a single country, PCD do better. German data by Kriesi et al. (1995) contain 676 events for the period, whereas the PCD have 1,177 events. The magnitude of the sources used by the European PCD is even more impressive in regards to the ratio of number of sources to countries. On average, in the European PCD one country is covered by about 16 news media (16:1) while in Banks' data (1991), the ratio drops to 0.006 sources per country.

Besides the increased number of events, the inclusion of more sources in PCD has improved the data on two other critically important dimensions. First, PCD are based on news sources with a wide range of interests. The specialty of sources ranges from general to very specialized interests. For example, news services such as *Agence France Press*, *Associated Press*, and *Reuters* provide general news, while news media such as *Energy Economist* focuses on a theme, and thus offers detailed reporting on their specific issue. The ideological mixture also increases the diversity of the PCD. There are conservative news media sources such as *Die Welt* of Germany as well as progressive media sources such as *Greenpeace Business*. By using such variety of sources, PCD extend the breadth and depth of sources to draw samples. And, this makes PCD more useful and less vulnerable to bias because of individual sources.

Second, PCD include many local sources. Local sources such as *Belfast Telegraph*, *Cypriot Broadcasting Corporation*, *Neue Zürcher Zeitung* (Daily Zurich newspaper), and *Radio Telefis Eireann* (Radio and Television of Ireland) were rarely, if ever, used in coding protests in any of the traditional datasets. Using local sources is important, especially for protest data, because protests tend to be about local issues and last only for short periods of time. If an event is serious enough—usually tragic and deadly and/or about an internationally appealing issue—the event would be covered not only by local news media, but also by the general news media. However, such events only occur sporadically because protests are usually about mundane issues like wage increases,[8] and they often end with no injuries or arrests. Such events are not so interesting to an international audience—they usually have similar issues of their own—and therefore are not likely to be reported by general news sources. In other words, local sources will report events that global sources will not.

Local Sources

How important is it to use local sources in data collection? A precise measure may seem hard to obtain, but we can test the significance of using local sources by comparing the sample size of datasets. I hypothesize that the protests coded using local sources will significantly outnumber the protests coded from non-local sources. This hypothesis is tested by comparing datasets in two contexts: 1) in Europe, comparing traditional datasets which rarely include local sources to the European PCD, which uses local sources available in Lexis-Nexis; 2) in Asia, comparing the utility of Lexis-Nexis and Asian sources. The first comparison has already been found to support the hypothesis. As discussed, the European PCD use various European sources in Lexis-Nexis and have far more observations than the

This content downloaded from
34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

traditional datasets (Francisco 2006). From this, we find evidence supporting the hypothesis. The degree to which local sources have contributed to this achievement can be illustrated by the case of Switzerland, where 1,811 events were coded in the PCD from 1980 through 1995. Among them, 561 events—about 31%—were coded from local sources. Had PCD relied exclusively on American sources such as the *Associated Press* or the *New York Times*, 31% of the events might have been neglected.

The second test is done in the context of the two Asian countries, Burma and Korea, for which PCD are available (Burma coded by Ferrara 2003; Korea by Nam 2005). The PCD project has been expanded beyond Europe to Asia and Latin America[9] in order to answer intriguing questions about experienced protests and/or revolutionary changes. The expansion of the PCD project beyond the European context has brought a new challenge especially for Burma and Korea because Lexis-Nexis—one of the main sources of European PCD—does not include any Burmese or Korean sources. To compensate for this absence, coders of events in the two countries relied on historical accounts, neighboring country's news sources (*Bangkok Post*), and local newspapers available outside of Lexis-Nexis. This practice not only helped the coders collect more representative samples, but it also provided an opportunity to test the hypothesis laid out above. If the emphasis on the availability of local sources were a valid point, Lexis-Nexis alone might not be enough to collect optimal samples for these two countries. Therefore, I expect that data coded from only Lexis-Nexis will contain significantly fewer events than the currently available data. A brief examination of the Korean data will illustrate the significance of local sources.

Magnitude of Korean Data

Unlike the European PCD that used Lexis-Nexis as a main source of information, Korean data were collected using the KINDS[10] mainly because of the absence of Korean sources in Lexis-Nexis. KINDS, managed by the Korea Press Foundation, updates daily and allows researchers to search events using Boolean operators like "and" or "or" to select desired reports. In searching for protest and coercion, I conducted a pilot search to find appropriate words. I found two words, *pa-up* (strike) and *si-wi* (roughly translated to rally) to collect desirable samples. Because KINDS allows a search with synonyms, using the two words successfully led to finding a wide range of events.[11] The KINDS search engine uses as many as 10 daily national newspapers in Korea, which provides a relatively large number of sources for a country (10:1). The 10 newspapers include both major newspapers such as *Cho-Sun Daily*, *Dong-A Daily*, *Joong-Ang Daily* and relatively new and minor newspapers such as *Hankyoreh* and *Se-Ge Daily*. Ideological diversity is represented as well: while *Hankyoreh* is a leading liberal newspaper, newspapers such as *Cho-Sun Daily* and *Dong-A Daily* are on the opposite end of the ideological spectrum.

Table 2
Eight Most Commonly Observed Actions

1990		1991	
Events	Frequency	Events	Frequency
Strike	1,694	Occupation	1,522
Occupation	1,380	Strike	989
Rally	1,101	Rally	804
Lockout	316	Hunger strike	257
Demonstration	184	Obstruction	249
Hunger strike	174	Boycott	192
Riot	166	Lockout	192
Statement	135	Statement	129
Others	947	Others	781
Total events	6,097	Total events	5,115

Source: Korean PCD.

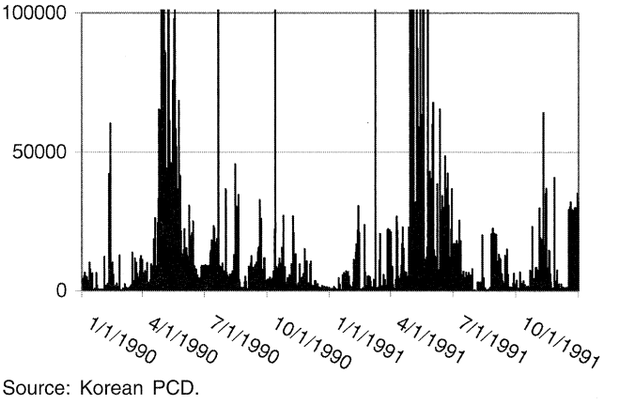
The collected data contain specific information and important theoretical utilities. Table 2 lists eight of the most frequently observed actions from 1990 to 1991. Accordingly, "Strike" and "Occupation" occurred most frequently, followed by "Rally." "Riot" subsided in 1991 while "Obstruction" became more popular. We also see that actions such as "Hunger strike," "Lockout," and "Statement" continued to occur persistently during these years. Table 2 shows there were 6,097 events, or 16.7 protest events per day, in Korea in 1990, and no single day went without an event. In comparison to other countries coded in PCD, this is by far the largest number of events observed in one year. Only events in Korea in the following year (5,115) approach this magnitude. Among the other countries, protests in the United Kingdom were second to Korea. In the U.K., there were 1,938 events in 1990, and 2,435 in 1991. This is about 5.31 events per day in 1990 and 6.67 events per day in 1991. This comparison demonstrates the significance of the magnitude of the Korean data.

The magnitude of protest activities can be illustrated by aggregating the total number of protesters[12] over a certain time period regardless of the type of action. Figure 1 shows vibrant dynamics of daily protesters[13] from 1990 through 1991. The other side of protest activity, coercion, is worthy of inspection. Coercion is a regime's coercive capacity to repress protesters,[14] and it can be measured by the daily sum of the number of protesters who are injured, arrested, or killed in a given time period. Figure 2 depicts the trend of daily coercion in Korea in 1990. Notice that types of coercion such as censorship or governmental regulation, which are not directly applied to protesters in action, are not measured. It is not because these instances of coercion are not important, but because the main aim of the data collection is to reveal the intimate and direct action dynamics between coercion and protesters.

Hypothesis Testing

How much of the extreme magnitude of Korean protests documented in the

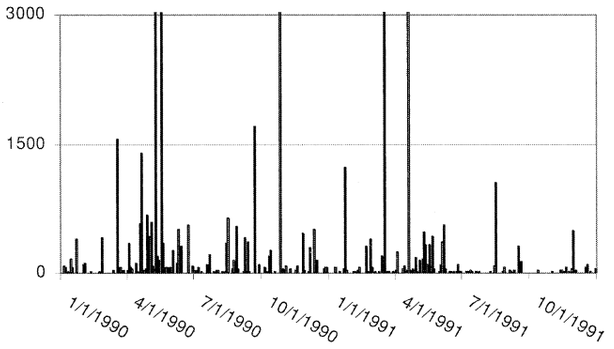
Figure 1
Protesters in South Korea 1990–1991 (truncated to show less than 100,000 protesters)



Source: Korean PCD.

This content downloaded from 34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

Figure 2
Repression in South Korea 1990–1991 (truncated to less than 3,000 repressed protesters)



Source: Korean PCD.

PCD is due to local data? To answer the question, I compared the Korean PCD to Korean data collected from Lexis-Nexis. As explained earlier, Lexis-Nexis provides invaluable variety in terms of sources. In the case of Korea, however, Lexis-Nexis lacks local sources. According to my previous argument, this lack of local Korean sources may lead to the compilation of unrepresentative and insufficient Korean data. On the other hand, the Korean PCD are based solely on local news sources. Therefore, a comparison between the two Korean datasets can answer whether it was the role of local sources that contributed to the huge and distinctive Korean dataset.

I hypothesized that protests coded from local source would outnumber protests coded from general sources. To test this hypothesis, I randomly sampled 10 weeks, out of the 104 weeks for two years covered by the Korean PCD (1990 to 1991), and compared the coverage—in terms of both the number of events and the number of protesters—of the Korean PCD to Lexis-Nexis.[15] Weekly protest activities were then matched for the 10 weeks. The same comparison was done with the Burmese PCD. Ferrara (2003) coded protests in Burma from March to October in 1988. Lexis-Nexis was used as a primary source; however, the *Bangkok Post* and other historical materials were also used to compensate for the absence of local sources. Therefore, I expect that the hypothesis will be valid in the case of the Burmese data as well. In other words, Ferrara's data are expected to include more events than what is collected from Lexis-Nexis alone.

The results confirm my *a priori* prediction. As Tables 3 and 4 show, using local data, or their equivalent, produced significantly richer datasets. In the case of the Korean data, the numbers of events in each and every week are different, and the number of events from Korean local sources is always larger than the number of events from the general sources of Lexis-Nexis. In Week 3, no events were reported by general sources, while Korean local sources reported 54 events. Similarly, in Weeks 40, 49, and 64, general sources did not report any events, but there were many events according to Korean sources. On average, there were 3.6 events per week reported by general sources, compared to 90.7 events per week reported by Korean sources, and a difference of means test shows the gap is statistically significant (see Table 3). Similarly, in weeks when no protest events were reported by general sources, many protesters participated in various events according to Korean sources. The average number of participants for the 10 weeks reflects a similar discrepancy between local (88,716.6) and general sources (21,391.5). According to a two-sample difference of means test, the number of protesters reported by Korean sources is also significantly larger than the number of protesters reported by general sources. The test on protesters shows that in terms of the number of protesters too, local sources are reporting significantly larger numbers.

The comparisons of the number of events and protesters in the two Burmese datasets validate the hypothesis as well (see Table 4). Even though the average number of events from local sources and from general sources was not statistically different, the mean number of protesters from local sources was significantly larger than the mean number of protesters from general sources.

Table 3
Number of Events and Protesters in Korea from 1990 to 1991: Comparing Different Sources

	Source of Data		Difference of Means Test (t)
	Local	General	
Mean Number of Events	90.7 (57.0)	3.6 (7.2)	5.4326**
Mean Number of Protesters	88,716.6 (171,063.5)	21,391.5 (52,174.1)	1.7834*

Source: Korean PCD and Lexis-Nexis.
Note: standard deviation in parentheses.
*p < 0.05; **p < 0.01 (one-tail test)

Table 4
Number of Events and Protesters in Burma in 1988: Comparing Difference Sources

	Source of Data		Difference of Means Test (t)
	Local	General	
Mean Number of Events	20.2 (20.6)	18.9 (16.4)	0.3163
Mean Number of Protesters	6,351,198 (7,623,406)	987,377.5 (1,108,271)	2.5097*

Source: Burmese PCD and Lexis-Nexis.
Note: standard deviation in parentheses.
*p < 0.05; **p < 0.01 (one-tail test)

Conclusion

Past studies of protests have used various data to enrich our knowledge of the topic. However, there have been problems with the data on which many studies rely. These problems include use of inconsistent categories, annually aggregated information, and reliance on a small number of sources. I argue that these problems

This content downloaded from 34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

have resulted in the neglect of valuable information and have left end-users with limited means to verify and manipulate information according to their needs. Moreover, the traditional datasets remain vulnerable to various types of bias due to their dependence on very few sources.

The PCD project solves these problems. By coding each and every reported protest event in a country, the project has eliminated the risk of losing detailed information and of imposing rigid categories. By relying on multiple sources, the project has adopted a measure to minimize bias. Therefore, these datasets have alleviated the various risks from which previous datasets have suffered. More important, the PCD include local coverage of protest events in sampled countries as important sources. In the case of the European PCD, Lexis-Nexis provided an excellent opportunity to use the local coverage, contributing to the unprecedented magnitude of European PCD. Burmese and Korean data included local sources outside of Lexis-Nexis and benefited remarkably from their use. The importance of local sources in coding protest events is supported by statistical tests. The comparison of data indicates that datasets vary widely in terms of the information that they contain. The presence of such a huge discrepancy among datasets unequivocally affects research results. Any judgment on the quality of a dataset is open to debate; however, this study has demonstrated that the data sources one chooses matter, as does the manner of coding.

Notes

*I give my sincere and deep thanks to Erik S. Herron and Ronald A. Francisco for reviewing drafts of this paper and giving me indispensable comments and continuous encouragements. I would also like to thank Elizabeth Collins and Omur Yilmaz who edited drafts of this manuscript. My appreciation also goes to anonymous reviews of this paper and the editors of *PS: Political Science and Politics*. I would like to thank my late farther, Ki-Dae whose love and support were simply endless. I'll share any compliment with them, but only I will take responsibility for any error.

1. Such as the Correlates of War Project (Singer).

2. Without any justification such an assumption ignores the occurrence of progovernment riots such as those in Venezuela under the Chavez regime.

3. Ferrara (2003) studied different reactions in the Burmese military regime toward mass protests in summer 1988.

4. Their data contain events of France, Germany, the Netherlands, and Switzerland from 1975 to 1989.

5. For this point and for further discussion about using multiple sources, see Francisco (2006).

6. As of June 14, 2005, 27 European states from 1980 through 1995, two Asian states (Burma 1988 and South Korea 1990–1991), three Latin American states (Colombia 1988–1997, El Salvador 1979–1991, and Peru 1980–1995) are coded. European PCD (Francisco) and links to other datasets are available at http://lark.cc.ku.edu/~ronfran/data/index.html. See Appendix A for definitions of exemplary actions coded in the datasets.

7. Protesters are those who challenge state authority using various means, including but not limited to demonstrations, strikes, riots, and statements.

8. For example, about 26.9% (1,639 out of 6,097) of events in South Korea in 1990 and about 64.6% (673 out of 1,042) of events in Finland from 1980 to 1994 are directly related to the issue of wage increase (Korean and European PCD, respectively).

9. Burma 1988 was coded by Ferrara; South Korea 1990–1991, by Nam; and Colombia 1988–1997, El Salvador 1979–1991, and Peru 1980–1995, by Garrison.

10. KINDS at www.kinds.or.kr.

11. They include various words like *demo, nong-sung, hang-ui nong-sung, hang-ui demo, si-wi, hang-ui siwi, demonstration, si-wi woondong, jip-dan si-wi, jip-dan nong-sung, pa-up, no-jo pa-up*, and *no-dong jo-hap pa-up*.

12. The measurement is distinguished from others that gauge the number of events (Gupta, Singh, and Sprague 1993, 309). Counting the number of protesters offers a more accurate picture of protest activities because it can compare the significance of one event with 10,000 participants with 10 events each carried out by 10 participants.

13. Such daily measurements are often more desirable than annual or monthly analysis. Daily aggregation is intended to reveal dynamics with as much detail as possible, following Lichbach's advice that "to capture the dynamics of protest over time, ... studies should use as short a temporal aggregation as possible" (Lichbach 1985, 601). Thus, for example, the significance of a reaction (i.e., backlash) responding to an event a few days prior can be accounted for by daily data, unlike weekly aggregated data incapable of the same task.

14. Coercion can be measured in various ways. One popular scheme is to produce an ordinal variable. For Gurr and Moore (1997), for example, coercion is comprised of six ordinary scales counting events from relatively light repression such as conventional law enforcement to brutal repression such as genocide (1,088). However, ordinal variables usually contain limited information (Blalock 1979, 16–17).

15. See Appendix B for the detailed information about the numbers of events and protesters during the selected weeks and for their comparisons between local and general sources.

References

Banks, Arthur S. 1972. "Patterns of Domestic Conflict, 1919–1939 and 1946–1966." *Journal of Conflict Resolution* 16 (1): 41–50.

———. 1991. *Domestic Conflict Behavior, 1919–1966* [Computer file]. 2nd ICPSR ed. Compiled by Arthur S. Banks, State University of New York at Binghamton, Center for Comparative Political Research. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

———. 2001. *Cross-National Time Series, 1815–1999*. Binghamton, NY: Center for Comparative Political Research, State University of New York.

Blalock, Jr., Hubert M. 1979. *Social Statistics, Revised Second Edition*. New York: McCraw-Hill.

Ferrara, Federico. 2003. "Why Regimes Create Disorder: Hobbes's Dilemma during a Rangoon Summer." *Journal of Conflict Resolution* 47 (3): 302–25.

Francisco, Ronald A. 2006. "Events coding: Protest and Repression." Under Review.

———. *European Protest and Coercion Data.* http://lark.cc.ku.edu/~ronfran/data/index.html.

Gupta, Dipak K., Harinder Singh, and Tom Sprague. 1993. "Government Coercion of Dissidents: Deterrence or Provocation?" *Journal of Conflict Resolution* 37 (2): 301.

Gurr, T. Robert, and Will H. Moore. 1997. "Ethnopolitical Rebellion: A Cross-Sectional Analysis of the 1980s with Risk Assessments for the 1990s." *American Journal of Political Science* 41: 1079–1103.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Kriesi, Hanspeter, Ruud Koopmans, Jan Willem Dyvendak, and Macro G. Guigni. 1995. *New Social Movements in Western Europe: A Comparative Analysis*. Minneapolis: University of Minnesota Press.

Lichbach, Mark I. 1985. "Protest: Random or Contagious? The Postwar United Kingdom." *Armed Forces and Society* 11 (4): 581.

Nam, Taehyun. 2005. *Korean Protest and Coercion Data, 1990–1991*. www. poliscience.org.

Singer, J. David. *The Correlates of War Project.* University of Michigan. www.umich.edu/~cowproj/dataset.html#MID.

Taylor, Charles Lewis, and David A. Jodice. 1983. *World Handbook of Political and Social Indicators*. New Haven: Yale University Press.

Woolley, John T. 2000. "Using Media-Based Data in Studies of Politics." *American Journal of Political Science* 44 (1): 156–173.

This content downloaded from 34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms

Appendix A
Definitions of Action

Action	Definition
Strike	Withholding labor for economic or political reasons
Occupation	Occupying state buildings or workplaces
Rally	Stationary protest gathering
Demonstration	Protest gathering that generally moves short distances or focuses on a specific target
Hunger strike	Refusal to accept food for political reasons; forced state feeding is reported
Riot	Generally used for violent prison protests, but also applies to purposeful violence against police
Statement	A group makes a public statement
Boycott	Politically based refusal to act, buy, or cooperate

Source: Codebook available at http://lark.cc.ku.edu/~ronfran/data/index.html.

Appendix B
Number of Events and Protesters

Number of Events and Protesters: Comparing Different Sources

Korea from 1990 to 1991					
Week	Date	Events from Local Sources	Events from General Sources	Protesters from Local Sources	Protesters from General Sources
3	Jan. 15–21, '90	54	0	13,419	0
10	March 5–11	41	1	6,173	8
19	May 7–13	235	23	563,749	167,447
40	Oct. 1–7	94	0	28,885	0
45	Nov. 5–11	104	7	78,892	26,635
49	Dec. 3–9	115	0	27,617	0
50	Dec. 10–16	93	3	16,109	1,803
55	Jan. 15–21, '91	53	1	11,879	21
64	March 19–25	73	0	15,650	0
87	Aug. 27–Sep. 2	45	1	124,793	18,001
Mean		90.7	3.6	88,716.6	21,391.5
Standard Deviation		57.0	7.2	171,063.5	52,174.1

Burma in 1988					
Week	Date	Events from Local Sources	Events from General Sources	Protesters from Local Sources	Protesters from General Sources
1	March 1–3	0	0	0	0
6	April 1–7	0	0	0	0
17	June 17–23	12	14	26,001	229,610
19	July 1–7	1	7	390	390
24	Aug. 5–11	36	36	6720,945	86,597
26	Aug. 19–25	42	38	20,311,412	2,360,385
27	Aug. 26–Sep. 1	57	21	18,743,046	2263,520
30	Sep. 16–22	36	47	7,704,840	2,725,600
31	Sep. 23–29	9	17	7,000,038	1,405,966
32	Sep. 30–Oct. 6	9	9	3,005,306	801,707
Mean		20.2	18.9	6,351,198	987,377.5
Standard Deviation		20.6	16.4	7,623,406	1,108,271

This content downloaded from
34.192.2.131 on Thu, 13 Apr 2023 15:59:45 UTC
All use subject to https://about.jstor.org/terms